A Path-Distance Model Explains the Structure of the Genetic Code

Brian K. Davis :: Research Foundation of Southern California, Inc., La Jolla, California :: davis@resfdnsca.org

Brian K. Davis, D.Sc.
Research Foundation of
 Southern California, Inc.
P.O. Box 13595
La Jolla, CA 92039
Tel./Fax 858 535 0899
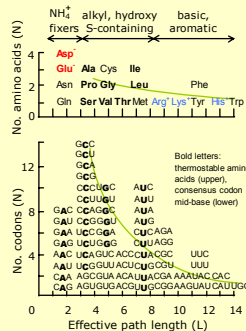http:www.resfdnsca.org

Introduction

The genetic code of the last common ancestor (LCA), or a minor variant of it, is present in all species. Its origin, in the pre-LCA era, has remained an enigma at the core of Molecular Biology for four decades. My analysis reveals that the diverse regularities observed in code structure correlate strongly with path-distances in amino acid synthesis. This clearly indicates that the code evolved by adding amino acids as they appeared, during the growth of synthesis pathways outward from central metabolism.

Objectives

- Establish whether the path-distance model of code evolution provides a general explanation for code structure.
- Identify the mechanism responsible for coordinating code evolution with the growth of amino acid synthesis pathways.

Design of Study

- Specify amino acid path-distances as number of reaction steps for synthesis, measured from the citrate cycle.
- Match each regularity in code structure with amino acid path-distances.
- Evaluate phylogenetic kinship between the conserved trace of pre-LCA tRNA species, to seek evidence they were cofactors in early amino acid synthesis.

Synthesis Pathways Order Amino Acids and Codons

Amino acids with short, medium and long paths are chemically distinct and encoded differently.

(1) Four NH_4^+ fixers have codons solely from the NAN column, both anionic residues (red) and 1-2 step paths.

(2) Ten amino acids have alkyl, hydroxy or S-bearing side-chains, and 4-7 step paths. Consensus NCN, NGN and NUN triplets code for 4-, 5- and 7-step residues, respectively.

(3) Six basic (blue) and aromatic residues form on 9-14 step paths and are encoded mostly by codon doublets.

A 14-fold, exponential fall-off in codon assignments accompanied path extension from 4 to 14 steps, in a gradual freezing of the code.

Amino Acid Path-Distance Patterns in the Genetic Code

- Codons for amino acids with 2-, 4-, 5-, and 7-step paths exhibit a columnwise distribution: p = 5.6x10⁻¹⁰
  NAN→NCN→NGN→NUN

- Codons in 7 of 8 intact boxes for group 1+2 amino acids (2-7 step). Codons in 6 of 8 subdivided boxes for group 3 residues (9-14 step). Codon distribution (NCN = NUN):
  p(1+2 vs. 3) = 4.6x10⁻⁵

Code Regularities Correlate with Amino Acid Path-Distances

- A and B show allocation of stable intact code boxes coincided with assignment of G/C enriched triplets to small amino acids (4-6 step paths) in code formation.
- C, decline in C oxidation number with path-distance in residues from citrate cycle.
- D, encoding of polar and hydrophobic residues by NAN and NUN codons, respectively, conforms with columnwise growth at stage 2 and 7 of code formation.
- E, triplets accumulate fastest in code boxes and slowest as single codons, consistent with subdivision of boxes into doublets, then doublets into singles.
- F, elevated mid-base coding capacity agrees with most (ten) amino acids entering the code during columnwise expansion through codon mid-base substitutions.

Monophyletic Pre-LCA tRNA bear Sibling Amino Acids

- Amino acid source: = α-KG, = Pyr, = OA, = PG, = Shi
- Superscript, amino acid path-distance
- [A], pre-LCA tRNA core group
- AUG, anticodon triplet (3' to 5')
- 3, cluster root in tree
- 14/18, RNA for synthetically related amino acids to total number in cluster

Conserved traces of pre-LCA tRNA sequences in tree show early tRNA adaptors for amino acids, derived from a common precursor, diversified from a common ancestor (sibling adaptor) and were cognate for similar codons.

- Positive tRNA Jukes-Cantor distance vs. amino acid path-distance: p(β ≤ 0) = 5.2x10⁻².
- Amino acid family stacking order: p = 2.05x10⁻².
- tRNA amino acid identity in tree clusters: p = 1.36x10⁻⁸.
- Anticodon 3'-base homology in clusters: p = 1.04x10⁻⁵.

Phases Identified in Code Evolution

- Amino acids with short (group 1), medium (group 2) and long (group 3) paths were encoded in distinct phases of code evolution: NH_4^+ Fixers, Expansion, Overprinting.
- Domains of contiguous codons read by related tRNA, with core group homology and bearing sibling amino acids, spread along code rows during its formation.

Conclusions

- Reconstruction of the path of code evolution that led to its diverse structural regularities has been achieved, upon equating the temporal order of amino acid addition to the code with the number of steps required for synthesis. The features of code structure unified by the path-distance model include,
  - Taylor-Coates amino acid size/code box rule
  - Dunnill G/C codon base/box rule
  - Dillon reductive amino acid synthesis prediction
  - Woese amino acid clusters
  - Dillon codon set subdivision rule
  - Perlwitz codon-base coding capacity ratio
  together with sixteen other regularities.

- Bifunctional pre-LCA tRNA species, serving as cofactors in amino acid synthesis and adaptors in translation, were credited with coordinating code formation with the growth of amino acid synthesis pathways.

- Frozen within the Standard Code are vestiges of previous codes. They provide compelling evidence for a chemoautotrophic origin of life on a cationic mineral surface:
  - Residues in the first proteins were traced to a primal NH_4^+ fixing mechanism coupled to the citrate cycle, which fixes CO_2, autocatalytically, under reductive early Earth conditions.
  - Early proteins were polyanionic and they became increasingly hydrophobic during code expansion, consistent with prolonging their dwell-time on a cationic mineral surface within an aqueous system.

- Extinction of the charge-attraction principle on completion of expansion phase, with subsequent incorporation of basic residues encoded by triplets captured in subdivided error-prone boxes, supports appearance of the first membrane-covered cells as the code began the final phase of its formation.

Reference

B. K. Davis, in Leading-Edge Messenger RNA Research Communications, M. H. Ostrovskiy, Ed. (Nova Science, New York, in press).